| | |
|---|---|
| **Iniziato** | martedì, 7 gennaio 2020, 15:05 |
| **Stato** | Completato |
| **Terminato** | martedì, 7 gennaio 2020, 15:31 |
| **Tempo impiegato** | 26 min. 17 secondi |
| **Punteggio** | 10,50/15,00 |
| **Valutazione** | **21,00** su un massimo di 30,00 (**70**%) |

Domanda **1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is different from the others?

**Scegli un'alternativa:**

a. Dbscan

b. SVM

c. Neural Network

d. Decision Tree

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following characteristic of data can reduce the effectiveness of K-Means?

**Scegli un'alternativa:**

a. Presence of outliers

b. All the variables have the same distribution of values

c. All the variables are the same range of values

d. Presence of values with high frequency

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Consider the transactional dataset below

| ID | Items |
|---|---|
| 1 | A,B,C |
| 2 | A,B,D |
| 3 | B,D,E |
| 4 | C,D |
| 5 | A,C,D,E |

Which is the *confidence* of the rule B ⇒ E?

**Scegli un'alternativa:**

a. 20%

b. 50%

c. 33%

d. 100%

Domanda **4**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

What is the *cross validation*

**Scegli un'alternativa:**

a. A technique to improve the speed of a classifier

b. A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set

c. A technique to improve the quality of a classifier

d. A technique to obtain a good estimation of the performance of a classifier with the training set

Domanda **5**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the effect of the *curse of dimensionality*

**Scegli un'alternativa:**

a. When the number of dimensions increases the classifiers cannot be correctly tuned

b. When the number of dimensions increases the computing power necessary to compute the distances becomes too high

c. When the number of dimensions increases the euclidean distance becomes less effective to discriminate between points in the space

d. When the number of dimensions increases the results tend to be prone to overfitting

Domanda **6**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

which of the formulas below computes the *recall* of a binary classifier?

**Scegli un'alternativa:**

a. (TP + TN) / (TP + FP + TN + FN)

b. TN / (TN + FP)

c. TP / (TP + FP)

d. TP / (TP + FN)

positives divided by

Domanda **7**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following types of data allows the use of the euclidean distance?

**Scegli un'alternativa:**

a. Document representations

b. Ordered data

c. Transactional data

d. Points in a vector space

Domanda **8**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

Which of the following is a base hypothesis for a bayesian classifier?

**Scegli un'alternativa:**

a. The attributes must be statistically independent inside each class

b. The attributes must have zero correlation

c. The attributes must be statistically independent *inside* each class, since the substitution of the joint p probabilities of the values is conditioned to a *class*

d. The attributes must have negative correlation

Domanda **9**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

In a decision tree, the number of objects in a node...

**Scegli un'alternativa:**

a. ...is not related to the number of objects in its ancestor

b. ...is smaller than or equal to the number of objects in its ancestor the decrease in size is at least one

c. ...is smaller than the number of objects in its ancestor

d. ...is bigger than the number of objects in its ancestor

Domanda **10**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

In data preprocessing, which of the following are the objectives of the *aggregation* of attributes

**Scegli una o più alternative:**

a. Reduce the variability of data

b. Obtain a more detailed description of data opposite direction

c. Obtain a less detailed scale

d. Reduce the number of attributes or distinct values

Domanda **11**

Parzialmente corretta

Punteggio ottenuto 0,50 su 1,00

Which of the following statements regarding the discovery of association rules is true? (One or more)

**Scegli una o più alternative:**

a. The confidence of a rule can be computed starting from the supports of itemsets

b. The support of a rule can be computed given the confidence of the rule

c. The confidence of an itemset is anti-monotonic with respect to the composition of the itemset

d. The support of an itemset is anti-monotonic with respect to the composition of the itemset

Domanda **12**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following is a strength of the clustering algorithm DBSCAN?

**Scegli una o più alternative:**

a. Ability to separate outliers from regular data

b. Requires to set the number of clusters as a parameter

c. Ability to find cluster with concavities

d. Very fast computation

Domanda **13**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the two binary vectors below, which is their similarity according to the Simple Matching Coefficient?

**a b c d e f g h i j**

1 0 0 0 1 0 1 1 0 1

1 0 1 1 1 0 1 0 1 0

**Scegli un'alternativa:**

a. 0.5

b. 0.3

c. 0.2

d. 0.1

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

After fitting DBSCAN with the default parameter values the results are: 0 clusters, 100% of noise points. Which will be your next trial?

**Scegli una o più alternative:**

a. Reduce the minimum number of objects in the neighborhood

b. Reduce the minimum number of objects in the neighborhood and the radius of the neighborhood

c. Decrease the radius of the neighborhood

d. Increase the radius of the neighborhood

Domanda **15**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

The *information gain* is used to

**Scegli un'alternativa:**

a. select the attribute which maximises, for a given test set, the ability to predict the class value

b. select the attribute which maximises, for a given training set, the ability to predict the class value

c. select the class with maximum probability

d. select the attribute which maximises, for a given training set, the ability to predict all the other attribute values